

1

확률분포함수 개념

1. 확률변수 random variable X

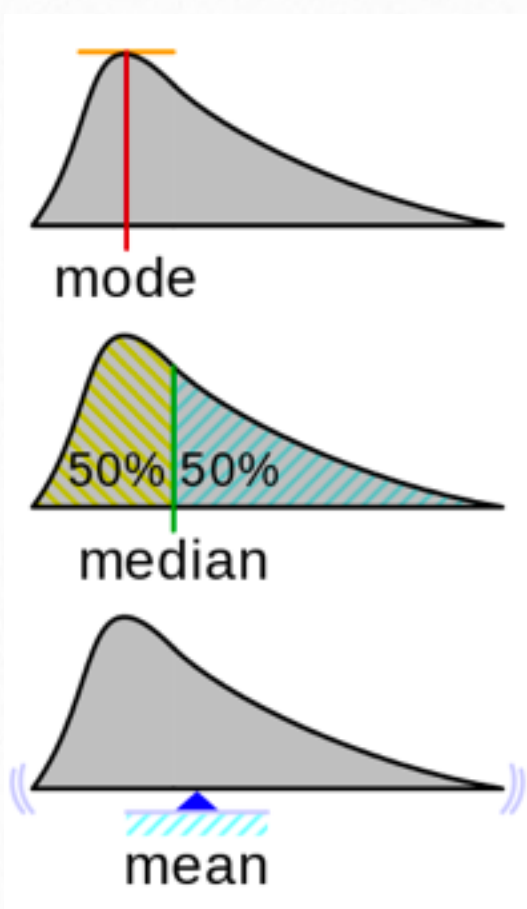
개체의 관심 특성(수능점수, 성별, 거주지 등)에 대한 관측값(데이터)

수학적 정의 : $X(\omega) = x$, ω 조사된 값 \Leftrightarrow 함수

2. 확률분포함수 prob. density function

$f(x)$: 확률변수 값이 input, 확률이 output인 함수

확률변수에 대한 모든 정보를 가지고 있음



1) 확률분포함수 성질

$$1) f(x) > 0$$

$$2) \int f(x) dx = 1$$

2) 기대값 expected value

$$\mu = E(x) = \int xf(x) dx : \text{평균 mean}$$

$$V(x) = E(X - \mu)^2 = \int (x - \mu)^2 f(x) dx : \text{분산}$$

Mode 최빈값 : 발생 가능성이 가장 큰 확률변수 값

$$\text{중앙값 median } \int_{-\infty}^m f(x) dx = 1/2$$

3. 활용예제

영역	성별	표준점수		등급(%)								
		평균	표준편차	1	2	3	4	5	6	7	8	9
국어	남	96.3	21.1	3.5	6.0	9.8	14.2	18.6	18.8	13.8	9.7	5.6
	여	100.4	18.6	3.1	6.1	11.7	18.0	22.5	18.9	10.9	6.0	2.8
	전체	98.3	20.0	3.3	6.0	10.7	16.1	20.6	18.9	12.3	7.8	4.2
수학가	남	98.9	20.2	6.2	7.5	9.4	17.6	17.3	17.4	13.1	8.7	2.9
	여	98.9	18.4	3.4	5.8	8.5	18.8	20.5	20.6	14.0	6.4	2.0
	전체	98.9	19.6	5.2	6.9	9.0	18.0	18.5	18.6	13.4	7.9	2.6
수학나	남	98.0	19.7	3.8	8.7	8.7	14.6	21.8	18.0	12.0	8.2	4.4
	여	99.8	19.1	3.3	9.5	10.4	16.3	21.9	18.1	11.0	6.2	3.3
	전체	99.0	19.4	3.5	9.1	9.7	15.5	21.8	18.0	11.4	7.1	3.8
영어	남	96.4	20.7	3.3	5.2	10.1	16.2	18.2	18.3	13.8	8.5	6.2
	여	99.8	18.7	3.6	5.7	11.4	18.8	20.4	18.5	12.7	6.1	2.8
	전체	98.1	19.8	3.5	5.5	10.8	17.5	19.3	18.4	13.2	7.3	4.5

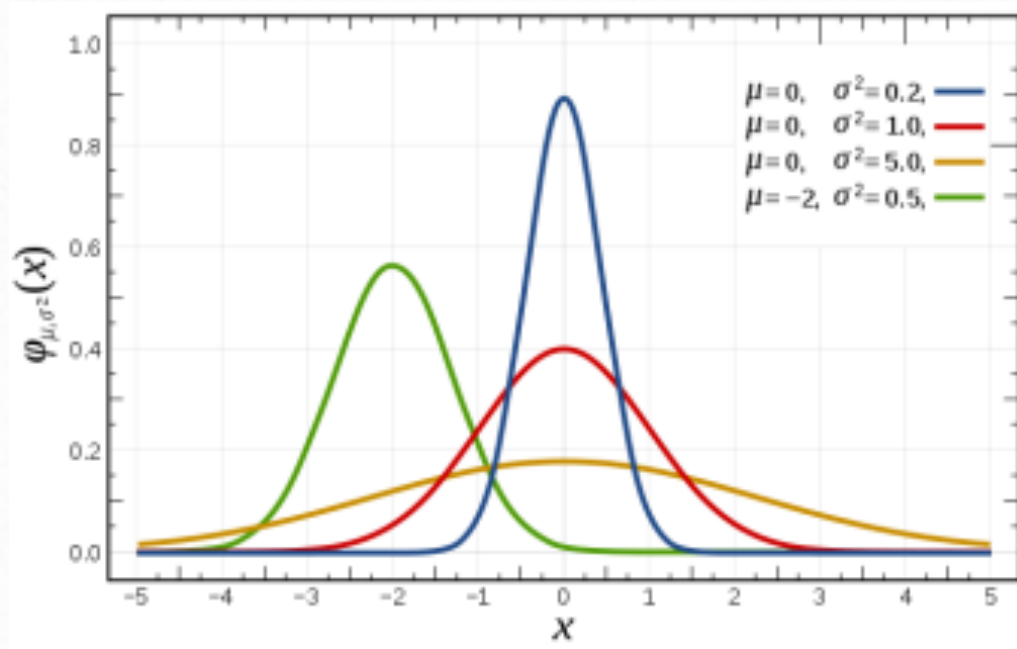
2017년 수능 성적

평균, 표준편차만 주어진 경우 - 전국 수능 응시자의 수능점수 정보를 알 수 있나?

아니다. 우리는 수능성적의 $f(x)$ 를 알 방법이 없다.

일단, 통계학에서 가장 중요한 분포인 정규분포라 가정하자. μ =평균, 표준편차= σ

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, -\infty < x < \infty$$



표준화

$$z = \frac{x - \mu}{\sigma} \sim f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}, -3 < z < 3$$

Mission

- 1) 수능국어점수 정규분포함수 (남자, 여자) 그리시오.
- 2) 확률분포함수의 성질을 가지고 있나?
- 3) 평균과 분산은 얼마인가?
- 4) 최빈값?
- 5) 수능 국어 1등급 상위 3.5%(3.1%) 점수를 구하시오.

알아두기

(1) $f(x) = x^n$ (미분) $f'(x) = \frac{df(x)}{dx} = nx^{n-1}$

(2) $f(x) = (ax^2 + \sqrt{x})^n$ (미분)
 $f'(x) = n(ax^2 + \sqrt{x})^{n-1}(2ax + 1/2x^{-1/2})$

(3) $f(x) = e^x$ (미분) $f'(x) = \frac{df(x)}{dx} = e^x$